
Hybrid Repeat/Multi-point Sampling for Highly Volatile Objective Functions

Brett W. Israelsen*

Department of Computer Science
University of Colorado
Boulder, CO 80309
brett.israelsen@colorado.edu

Nisar Ahmed†

Department of Aerospace Engineering Sciences
University of Colorado
Boulder, CO 80309
nisar.ahmed@colorado.edu

Abstract

A key drawback of the current generation of artificial decision-makers is that they do not adapt well to changes in unexpected situations. This paper addresses the situation in which an AI for aerial dog fighting, with tunable parameters that govern its behavior, will optimize behavior with respect to an objective function that must be evaluated and learned through simulations. Once this objective function has been modeled, the agent can then choose its desired behavior in different situations. Bayesian optimization with a Gaussian Process surrogate is used as the method for investigating the objective function. One key benefit is that during optimization the Gaussian Process learns a global estimate of the true objective function, with predicted outcomes and a statistical measure of confidence in areas that haven't been investigated yet. However, standard Bayesian optimization does not perform consistently or provide an accurate Gaussian Process surrogate function for highly volatile objective functions. We treat these problems by introducing a novel sampling technique called Hybrid Repeat/Multi-point Sampling. This technique gives the AI ability to learn optimum behaviors in a highly uncertain environment. More importantly, it not only improves the reliability of the optimization, but also creates a better model of the entire objective surface. With this improved model the agent is equipped to better adapt behaviors.

1 Introduction

Due to current and expected logistical and fiscal constraints the Department of Defense (DoD) has been focusing on simulation-based training of warfighters. To this end, the Not-So-Grand Challenge was developed, with the specific goal to investigate solutions for current and future simulation training systems. As part of this challenge different autonomous agents were developed and evaluated based on their ability to mimic a human pilot in given situations [1]. Even though an autonomous agent may mimic a human pilot there still remains the question of whether it can adapt based on the adversaries responses.

Previous related work focused on optimization of target allocation, tactics, and mission plans for aerial combat [2–5], but have not addressed adaptation of autonomous AI decision-makers with tunable behavioral parameters. This work specifically examines how an agent with tunable parameters that govern overall behavior can be adapted to optimize an objective function that quantifies engagement outcomes. Beyond optimizing some outcome metric, it is also important that the agent have a realistic representation of the entire objective function. This will allow the AI to anticipate the likelihood of successful engagements with adversaries (human or AI) under different uncertain conditions. These

*Graduate Researcher, corresponding author, <http://bisraelsen.github.io/>

†Assistant Professor, <http://www.cohrint.info/>

behavioral changes could be based on adapting to the adversary’s skill level (it is not desirable to use the same difficulty level for novice and advanced pilots), or the adversary’s ability to exploit a weakness of the agent (the objective function is changing).

There are several challenges that make optimization of AI behavioral parameters difficult in this application:

1. Simulating an engagement can be costly. Beyond the financial expense of operating the simulation environment, contributions to the cost may also include the involvement of skilled labor/participants with limited availability, and the wall-clock duration of the simulation itself.
2. The objective function to be optimized is not known a priori, and when sampled is generally nonlinear and noisy. Consequently, many traditional optimization methods are not applicable.
3. Due to the realistic nature of the simulations and the nature of aerial combat, the objective function is extremely volatile and uncertain (e.g. due to combined random effects of weather, terrain, sensor noise, psycho-motor time delays, etc.).
4. Besides only identifying the optimum performance, the agent should also try to obtain some model of the overall objective function. This can allow the agent to be adaptive and have some notion of what outcomes might arise when modifying behavior parameters, without having to exhaustively search over a high-dimensional parameter space. In addition, this model can also be used to generate a useful estimate of the expected performance of adversaries for a wide range of scenarios, using only a small number of test evaluations.

Gaussian Process based Bayesian optimization (GPBO) is well-suited for addressing points 1 and 2. However, we show that typical application of GPBO is not well suited to address points 3 and 4, in this application. We introduce and demonstrate a new sampling approach called Hybrid Repeat/Multi-point Sampling (HRMS) that yields some promising results in this regard, by capturing more statistical information about the objective function on each iteration of the optimization. With proper configuration, HRMS is able to not only identify the optima more reliably than standard GPBO, but also yields a more useful surrogate representation of the objective surface. Finally, it generally does this using no more total function evaluations than traditional GPBO.

2 Methodology

The problem is defined as an air combat scenario with autonomous red and blue force agents. Each of the agents has behavioral parameters given by the parameter vectors \mathbf{x}_r and \mathbf{x}_b respectively. The goal is to optimize an objective function $y_i(\mathbf{x}_r, \mathbf{x}_b)$, where $y_i(\cdot)$ must be evaluated using a high-fidelity combat simulation. For this work \mathbf{x}_r is constant, and the optimization will only be changing \mathbf{x}_b . The remainder of this paper uses the following: $y = y_{TTK}(\cdot)$ (TTK stands for time to kill), and $\mathbf{x}_b = \{x_1, x_2\} = \{\text{launch}, \text{intspeed}\}$, where *launch* is the time to launch weapon once lock is acquired and *intspeed* is the intercept speed (i.e. the speed at which the blue fighter moves into close the range on red once engaged). Note that there are 11 total behavior parameters available in the aerial-combat simulation, but we only investigate two here.

Figure 1 depicts the high-level learning process. Both the red and blue agents have tunable parameters, but only blue is being changed in this scenario. Figure 2 shows 1d slices of the objective function for $x_b = \text{intspeed}$ and $x_b = \text{launch}$.

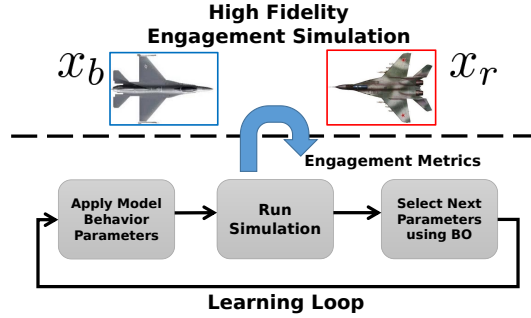


Figure 1: Learning Process Diagram. Representation of the engagement simulation environment (top) and the high-level learning loop (bottom)

Using an acquisition function $a(\cdot)$, there are two commonly used ways to search for the optimum. The first is single sampling (SS) where $y_i(\mathbf{x})$ is evaluated a single time at the $\arg \max_{\mathbf{x}} a(\cdot)$ of the acquisition function; this is the standard GPBO approach. The second is multiple (or batch) sampling (MS), in which $y_i(\mathbf{x})$ is evaluated at multiple different locations simultaneously. Finally, we propose a method called repeat sampling (RS), which is identical to SS except that the objective function will be evaluated repeatedly at the same location, $\arg \max_{\mathbf{x}} a(\cdot)$.

The intuitive reason for introducing repeat sampling is to obtain a more informative statistical sample of the objective function at every iteration. This is necessary because, for GPBO to work properly, the surrogate function needs to be a ‘sufficiently accurate’ representation of the true objective function. RS helps the GP to have more information regarding the noise of the true objective function, so that the GP can be a useful surrogate function in guiding the GPBO. This method is also used in traditional experimental design where it is called ‘replication’ (see [6] and [7, sec. 4.4.4.6]). Another consideration that makes repeat sampling or replication attractive is prohibitive cost to setting up new experiments, something that is not as much of a problem when dealing with computer simulations but becomes an important consideration with applied problems like training pilots.

The RS and MS strategies would be especially valuable when the objective function is less expensive to evaluate via simulation, and the experiments can be run in parallel without significantly increasing the overall cost of the optimization. In the following, $MS=3$ means that 3 batch samples will be selected. Likewise, $RS=3$ is where 3 samples will be taken at the same location. Note that SS is a special case where $RS=MS=1$. Finally, we refer to combined sampling where RS and MS are both greater than 1, as Hybrid Repeat/Multi-point Sampling (HRMS).

3 Experiments

Given a decision agent optimization problem, we perform experiments to investigate the performance of different acquisition functions for the aerial combat simulations. More importantly we wish to investigate the effect that varying RS and MS has on the optimization results.

Specifically, we evaluate three different, common, acquisition functions: Expected Improvement (EI) [8], upper confidence bound (GP-UCB) [9], and Thompson Sampling (TS) [10]. We also evaluate their corresponding batch sampling forms: q-EI [11], GP-UCB-PE [12], and multiple draws from TS. The different levels of RS and MS used are $RS = \{1, 3, 5, 10\}$ and $MS = \{1, 3, 5\}$. The GPML toolbox [13] is used for GP representation and hyperparameter inference. There is also an interface to the MSS air combat simulation engine made by Orbit Logic Inc. The kernel is the Matérn ARD kernel with $\nu = 3/2$ [14]. GPBO is run for approximately 500 function evaluations (approximately 500 because different HRMS configurations don’t allow exactly 500); 4 experiments are run per HRMS configuration using 20 random seed locations to bootstrap GP learning with MAP inference for hyperparameter optimization.

Figure 3 shows the estimated locations of the optimum as well as the optimum itself for the UCB acquisition function. The estimates for \mathbf{x} and y grow tighter together, and closer to the ground truth, as both RS and MS become greater than 1. The configurations marked by colored rectangles highlight that methods using solely SS, RS, and MS (shown in blue), underperform the method that combines both RS and MS greater than one (yellow). This finding is similar for the EI and TS functions as well. From this figure we can conclude that there are HRMS configurations that yield more repeatable optimization results than SS, and that neither RS or MS alone is clearly better.

On more detailed investigation of the results, for large RS the optimization tends to terminate early due to an ill-conditioned covariance matrix. This occurred because RS is ‘too big’ at those locations,

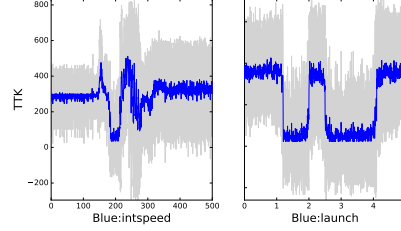


Figure 2: One-dimensional examples of TTK objective function, for $\mathbf{x} = \{\text{Blue:intspeed}, \text{Blue:launch}\}$. These figures were produced by holding all \mathbf{x}_r parameters constant, as well as all \mathbf{x}_b parameters except the one listed. The dark blue line represents the empirical μ and the shaded area is 2σ

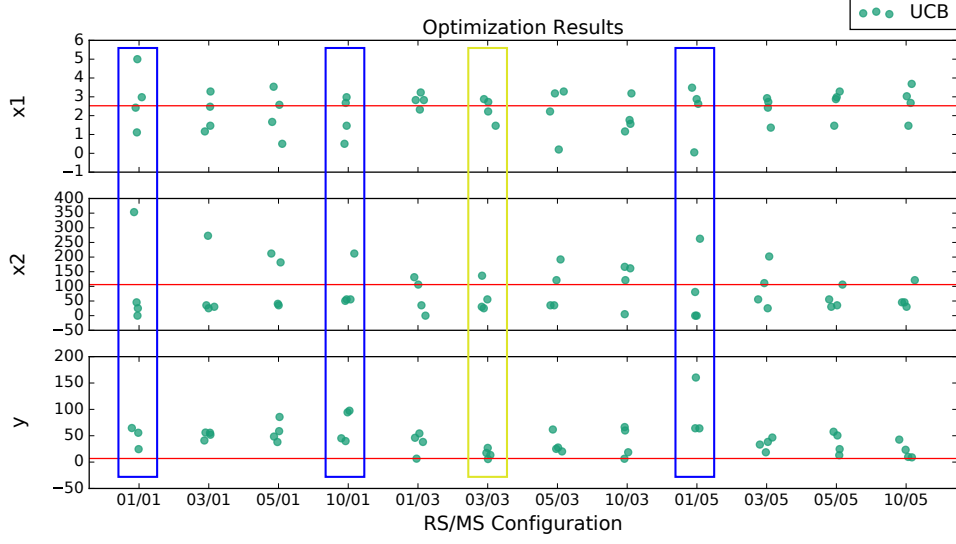


Figure 3: Scatter plots of the x_1 , x_2 and y values for different RS/MS configurations. Results from running GPBO for approximately 500 function evaluations. The red horizontal line is the ground truth value.

having returned too many nearly identical objective function values at the same (x_1, x_2) locations. The subsequent covariance matrix became too linearly dependent, which led to conditioning problems for GP inference. This suggests that there is clearly a trade-off between the benefits of RS and an unstable GP. We revisit this point in the conclusion section.

It is important to note that given a fixed time for optimization, in other words: not limiting the function evaluations for methods that perform more quickly, the total number of function evaluations in most cases does not exceed that of the SS strategy. This is mainly due to the overhead of calculating MS, and is illustrated in Figure 4, where only TS significantly exceeds the total function evaluations of SS.

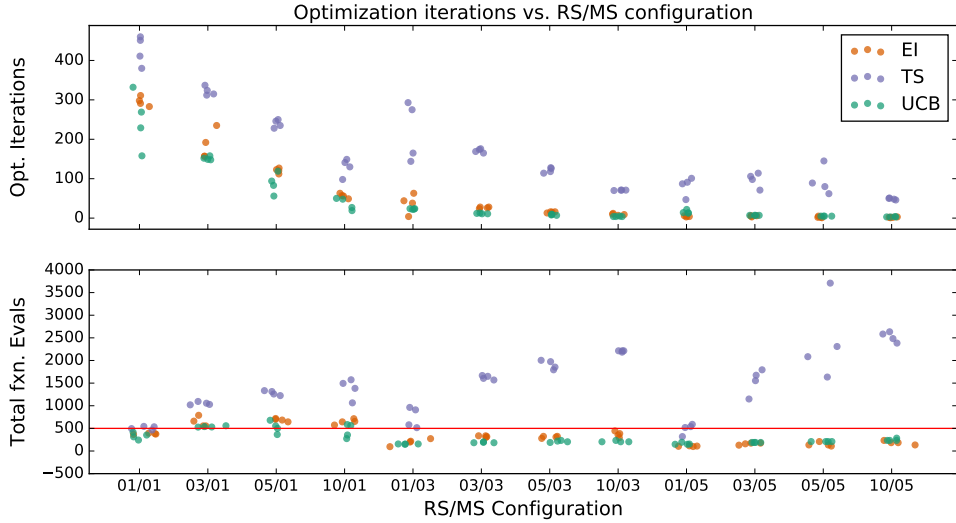


Figure 4: Plot showing the total amount of optimization iterations (Top), and the corresponding number of function evaluations (Bottom) after running each configuration for 2 hours. Note that, with the exception of TS, the total number of function evaluations for mixed RS/MS configurations generally doesn't exceed that of GPBO with SS

Figure 5 depicts some examples of the final GPs obtained during time limited optimization (again, allowing faster methods to use more evaluations/iterations) for three different HRMS configurations. The far left column is the ‘ground truth GP’ model that is obtained by training with several thousands of samples over the input space. The key insight is that the RS3/MS3 strategy yielded a GP that better represents the underlying stochastic function.

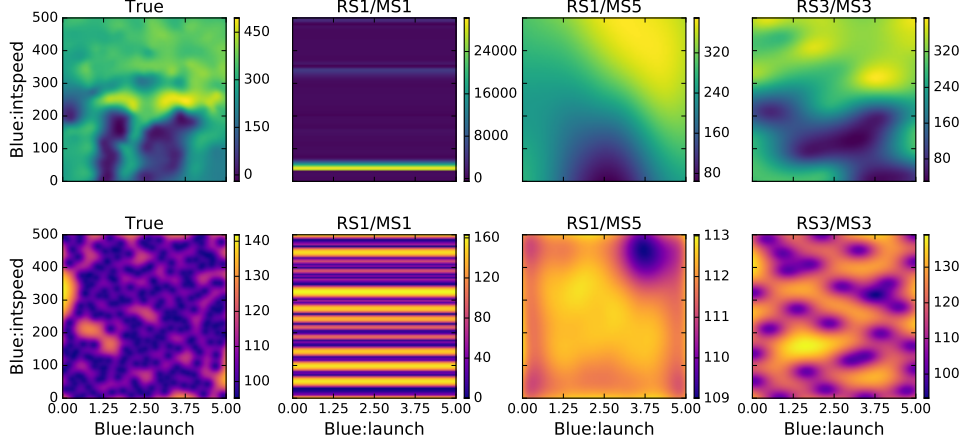


Figure 5: Table of figures illustrating the effect of combined RS/MS sampling using the UCB acquisition function. Top row is μ_{GP} bottom row is σ_{GP} . From left to right the first column is the truth surface obtained by high density sampling and fitting a GP to the data. The following columns show some example results from optimization runs using the indicated values for RS and MS. Each of the final 3 columns represents the optimization solution after 2 hours

These findings indicate that HRMS both improves the repeatability of the optimization, *and* the overall fidelity of the surrogate representation of the objective function. This applies for both a fixed computation time (i.e. not limiting faster methods like SS to have the same number of function evaluations), and number of function evaluations. These two phenomena are linked, i.e. the optimization is more repeatable (and reliable) because the surrogate representation is more accurate.

4 Conclusion

We have shown promising preliminary results of HRMS, a novel sampling strategy that helps improve both the repeatability/reliability of GPBO (a desirable feature for box optimization), and a better surrogate representation of the true objective surface. This surface can be used in understanding how the underlying process works and will allow an AI decision-maker to adapt in highly volatile and uncertain environments. Preliminary experiments show that improvements from HRMS are independent of the acquisition function for our application. They also show that the improved performance is due to the fact that adding both local and global information about the objective function at each time step makes the surrogate function more accurate. This accuracy yields a more efficient GPBO process, and does not require more function evaluations than the standard SS approach. Again, these results while promising still need to be examined more rigorously and be statistically verified.

To the best of our knowledge, HRMS has not been considered for GPBO previously. Further investigation regarding the relationship with replication, in design of experiments, needs to be explored more formally. There is still much work to be done regarding automatic selection of RS and MS. Perhaps only adding new data to the GP covariance function if it has sufficient variation might work, but would need to be formally assessed. Finally, the preliminary results shown here are being extended and verified in higher dimensional problems. The AI decision-maker in our aerial combat simulation, for instance, has 11 behavioral parameters that can be used for optimization. In higher dimensional spaces, the automatic selection criteria for RS and MS becomes more important as the surrogate function and optimization results become much more unwieldy and uncertain, and visual comparison is no longer available to verify the similarity between the true objective function and its surrogate representation.

Acknowledgments

We would like to acknowledge Kenneth Center, and Roderick Green with Orbit Logic Incorporated for their collaboration in designing the simulation framework used.

References

- [1] Margery J Doyle and Antoinette M Portrey. Rapid Adaptive Realistic Behavior Modeling is Viable for Use in Training. In *Proceedings of the 23rd Conference on Behavior Representation in Modeling and Simulation (BRIMS)*, 2014.
- [2] S. Mulgund, K. Harper, K. Krishnakumar, and G. Zacharias. Air combat tactics optimization using stochastic genetic algorithms. In *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.98CH36218)*, volume 4, pages 3136–3141, Oct 1998.
- [3] Sandeep Mulgund, Karen Harper, and Greg Zacharias. Large-Scale Air Combat Tactics Optimization Using Genetic Algorithms. *Journal of Guidance, Control, and Dynamics*, 24(1):140–142, Jan 2001.
- [4] Wen-Hai Wu and Qingdao Branch. Air Combat Decision-making for Cooperative Multiple Target Attack using Heuristic Adaptive Genetic Algorithm. In *Proceedings of the 4th International Conference on Machine Learning and Cybernetics*, volume 1, pages 473–478, Aug 2005.
- [5] P.G. Gonsalves and J.E. Burge. Software Toolkit for Optimizing Mission Plans (STOMP). In *Collection of Technical Papers - AIAA 1st Intelligent Systems Technical Conference*, volume 1, pages 391–399. American Institute of Aeronautics and Astronautics, Sep 2004.
- [6] Thomas Pyzdek and Paul A Keller. *Quality engineering handbook*. CRC Press, 2003.
- [7] Carroll Croarkin and Paul Tobian. e-Handbook of Statistical Methods. *NIST/SEMATECH*, Available online: <http://www.itl.nist.gov/div898/handbook>, 2016.
- [8] Donald R. Jones, Matthias Schonlau, and J William. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4):455–492, 1998.
- [9] Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. Jun 2010.
- [10] William R. Thompson. On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples. *Biometrika*, 25(3/4):285, Dec 1933.
- [11] Jialei Wang, Scott C. Clark, Eric Liu, and Peter I. Frazier. Parallel Bayesian Global Optimization of Expensive Functions. Feb 2016.
- [12] Emile Contal, David Buffoni, Alexandre Robicquet, and Nicolas Vayatis. Parallel Gaussian process optimization with upper confidence bound and pure exploration. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 8188 LNAI, pages 225–240. Springer, 2013.
- [13] Carl Edward Rasmussen, Christopher K I Williams, and ebrary Inc. GPML Toolbox, Dec 2006.
- [14] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. Adaptive computation and machine learning. MIT Press, Cambridge, Mass, 2006.